Green Energy: Advancing Bio-Hydrogen

Developing a model of metabolism linked to H_2 production in green algae

David Alber

Scientific Computing Center
National Renewable Energy Laboratory

NREL/PR-530-41988

Presented at the Workshop on Petascale Architectures and Performance Strategies held in Snowbird, Utah on July 23-26, 2007.

Project Participants

- NREL Basic Sciences: Michael Seibert (PI)
- NREL Scientific Computing Center: David Alber, Christopher Chang, Peter Graf, Wesley Jones (co-PI), and Kwiseon Kim (co-PI)
- Colorado School of Mines: Glen Murray (co-PI) and Matthew Posewitz (co-PI)
- Stanford University: Arthur Grossman (co-PI)

Project Overview

Goals

- Computationally model complete metabolism of green alga *Chlamydomonas reinhardtii*
- Develop tools for parameter discovery and optimization at organism level
- Advance knowledge of hydrogen-producing photosynthetic organisms

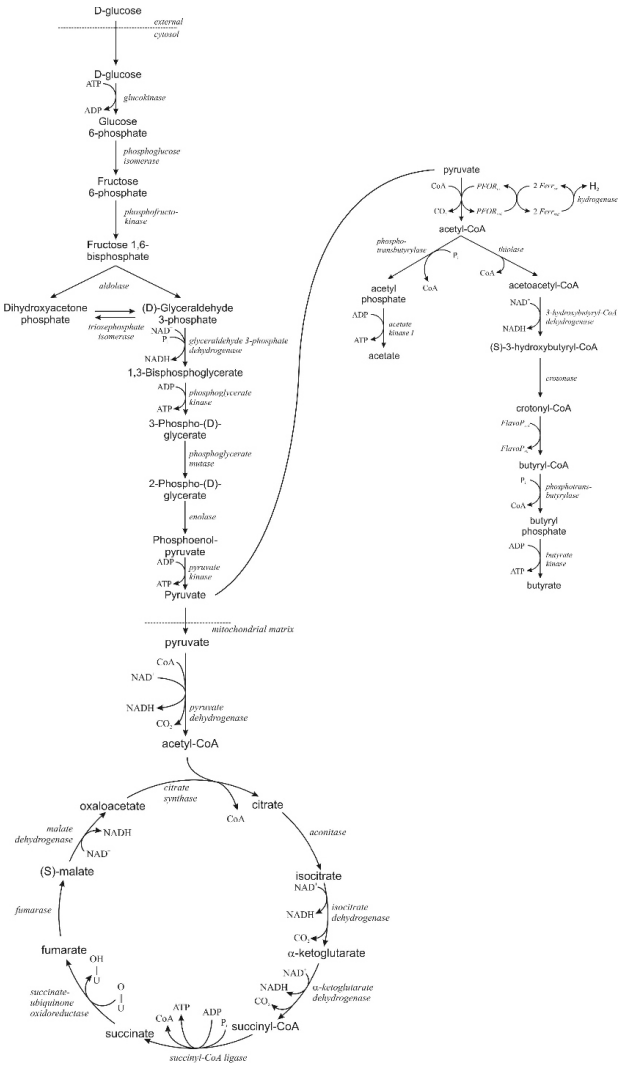
- Computational research part of larger project
- Funding through SciDAC (OASCR and OBER)
- Funding commenced six months ago

Science Background

- Create model of metabolic pathways with ODE represented by edges:

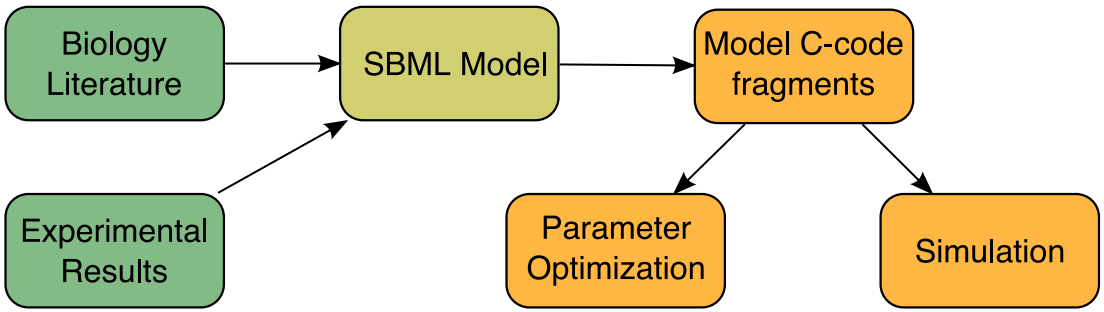
$$\frac{d[y_i]}{dt} = \frac{k_{\text{cat}} \cdot [E]_{\text{tot}} \cdot [y_j]}{[y_j] + K_M}$$

- k_{cat} and K_M: kinetic constant parameters being sought
- Some parameters well known experimentally, others not
- Employ optimization to determine parameter values for model



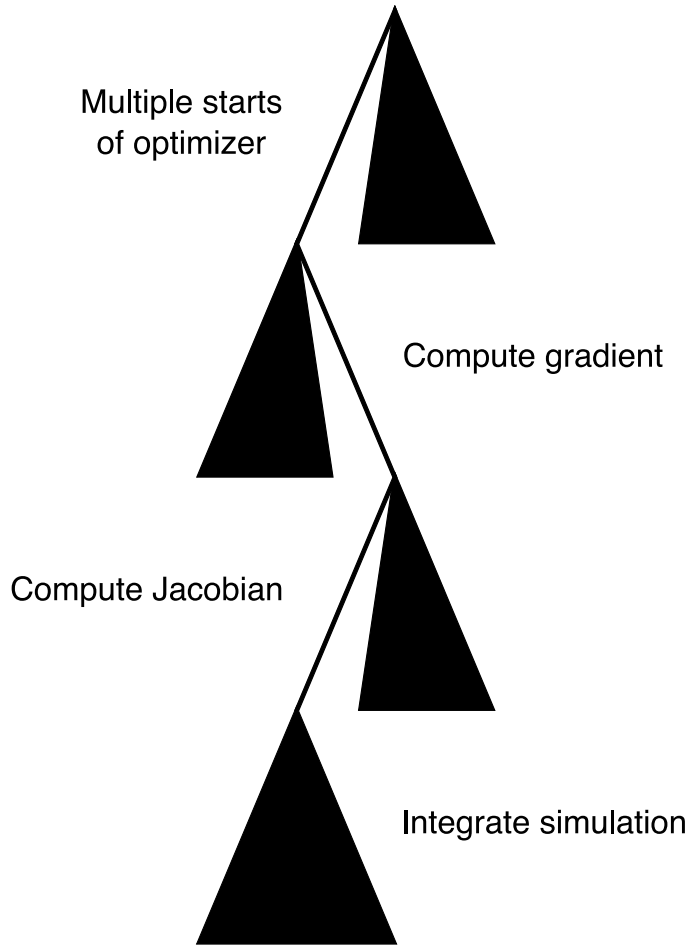
Science Background

- Primary objective is determination of “true” set of parameters
- Find parameters based on:
  - Species concentration targets ($y_i = \text{target at } t_{\max}$)
  - Lifelike networks (all pathways remain “alive”)
  - Species concentration insensitive to perturbations in parameters ($|\partial y_i / \partial k_j|$ small)



Parallel Programming Model

- Code in C, some Python
- Standard build tools
- Future design still flexible
- Hierarchical parallelism



Computational Methods

- Tools:
  - Systems biology tools:
    - Systems Biology Markup Language (SBML) – encodes the model
    - libSBML – interface for manipulating SBML
    - Systems Biology ODE Library (SOSlib) – produces ODEs from the model
  - ODE tools:
    - CVODES – solver for systems of ODEs, includes sensitivity analysis
  - Optimization tools:
    - Toolkit for Advanced Optimization (TAO)
- Status: early development and open to different external packages

I/O Patterns

- Input relatively small and limited to root processes
- Output small per optimization job
- For sampling parameter space, possibly large number of output files requiring additional level of processing

Visualization and Analysis

- Sample of questions to answer with visualization and analysis:
  - Why did the organism develop to use one set of parameters over some other set?
  - How easily satisfied are our constraints? How prevalent are local minima?
  - Which reactions are strongly coupled?
- Current methods visualize species concentration over time of simulation
- Future visualization:
  - Visualization to dynamically change subset of data being examined
  - Zoomable to traverse graph of model
  - Must be useful and intuitive to biologists

Tools

- Debuggers: gdb, Valgrind, perhaps TotalView in future
- <oXygen/> for editing model
- Further down the road:
  - Documentation (DocBook, LaTeX, Doxygen, something else?)
  - Test suite automation
  - Tighter coupling of visualization tools

Roadmap
Next two years

- Expand *C. reinhardtii* metabolic model (number of species in model will increase 5-10 times)
- Continue developing understanding of model and properties
- Build code and solve parameter search problems for current model
- Develop parallel code for larger models
- Incorporation of feedback from experimentalists to expand target concentrations list